Proceedings of The Institute of Acoustics

GLOTTAL LOSSES IN A COMPOSITE MODEL OF SPEECH PRODUCTION

C.SCULLY and E.ALLWOOD

DEPARTMENT OF LINGUISTICS AND PHONETICS
UNIVERSITY OF LEEDS

LOSSES IN SPEECH PRODUCTION

One way in which the acoustic block of an articulatory synthesiser can be improved is by better modelling of losses. Energy is lost from the vocal tract by the following processes:

radiation of sound from the mouth and nostrils, also radiation from the walls;
vibration of the yielding walls of the head, neck and chest;
heat conduction through the walls;
viscous losses at the walls;
turbulent as well as viscous losses in the glottis; also in severe constrictions of the vocal tract;
losses in the subglottal airways.

When the vocal tract acoustic system is modelled as one-dimensional waves on electrical transmission lines, viscous losses are represented as series resistances, heat conduction as shunt resistances and wall vibration and radiation as shunt impedances. Mouth radiation impedance is represented by appropriate termination of the line. Glottal losses can be related to a vocal fold model (1). The method of reflection coefficients (2), used by us, is faster to compute but assumes that pressure and volume velocity remain in phase. It does not model the formant frequency shifts and losses due to radiation impedance and losses within each vocal tract section; however, methods of doing so have been proposed (3). Many of the loss parameters are frequency dependent, but a single value for each must be selected in time-domain simulations. This limitation applies whichever method is employed in an articulatory synthesiser.

Our modelling has demonstrated that some, at least, of these losses must be simulated in order to achieve formant bandwidths similar to those of real speech. Viscous and heat losses may reasonably be neglected in a simplified acoustic model. They are apparently small compared with other types of losses, being much less than glottal and wall losses at low frequencies and much less than glottal and radiation losses at high frequencies (1). Radiation loss is proportional to frequency squared up to about 3 kHz for vowels other than those with a very large lip outlet area. Increased radiation loss should lower formant frequencies. The main effect of wall losses is seen at low frequencies: increased loss should raise first formant frequencies; the effect is expected to be noticeable in [ɑ]-type vowels especially. Increased glottal losses should raise formant frequencies. Increased losses during large glottis - occluded vocal tract portions were found to be essential; otherwise, the pressure wave trapped in the vocal tract took too long to decay during a simulated [s]. It seemed worthwhile to attempt a fairly sophisticated representation of glottal losses.

LARYNGEAL PROCESSES

The larynx is an important element of the articulatory, aerodynamic and acoustic blocks of the model (4). Consistency across these stages of speech production is sought. Within each block the aim is to simulate different speaker types while conforming to the physical constraints of the system. The larynx is assumed to contain two independent articulatory actions: abduction and adduction of the vocal folds for the control of glottal area $\overline{Ag}$; tensing of the vocal folds as a factor, called $\overline{Q}$, in fundamental frequency control. Aerodynamic conditions at the glottis are determined by the changing state of the whole respiratory system. Subglottal air pressure $\overline{Psg}$, oral air pressure $\overline{Pc}$ and volume velocity of airflow through the glottis $\overline{Ug}$ are obtained. Here, a variable $\overline{Vx}$ is the slowly changing 'd-c' component of the total, while Vx' is the rapidly changing acoustic 'a-c' component and Vx is the total time varying function. Ug' is the voice source in the acoustic block. It is the short circuit volume velocity, not the true volume velocity dependent upon the vocal tract filter function. Ug' is derived from a functional model of the larynx. Its waveform parameters are those of Fant (5). Parameter values depend on three controlling variables: transglottal pressure drop ($\overline{Psg} - \overline{Pc}$), glottal area $\overline{Ag}$ and vocal fold tension $\overline{Q}$.

It is predicted that the glottal contribution to formant bandwidths is approximated by $BW1(gl) = \dfrac{c^2 . Ag^2}{2.\pi . k . Ug . Vc}$ assuming a Helmholtz resonator, or twice this, for all formants, if the vocal tract is modelled as a tube (6). Vc is the volume of the cavity above the glottis and k (= 0.875) is an empirical constant. c is the velocity of sound in the vocal tract. The expression for glottal differential signal resistance assumes that turbulent losses dominate, although viscous losses are included also in the aerodynamic block of the model.

In the reflected pressure wave model of vocal tract acoustics, glottal losses are included by making the reflection coefficient RCg, at the glottal end of the first section, less than 1. Similarly the reflection coefficient RCm at the mouth termination is made greater than -1 ($|RCm| < 1$) for simulation of radiation losses. Increased glottal area $\overline{Ag}$ results in increased losses at the glottis since RCg is a function of $\overline{Ag}$, viz:- $RCg = \dfrac{a}{(\overline{Ag})^R}$

a and R are constants chosen such that at Ag = 0.05 ('normal' phonation value), RCg = RCg1 and at Ag = 0.15 ('breathy' voicing), RCg = RCg2. RCg1 and RCg2 are parameters whose values may be changed, but which are generally 0.8 and 0.5 respectively. An upper limit for RCg is specified, usually 0.8, as a representation of glottal losses during a voicing cycle when the vocal folds are tightly adducted.

A STUDY OF BANDWIDTHS USING THE MODEL

To investigate the influence of cavity volume Vc upon losses in the model two vocal tract shapes, as shown in Figure 1, were compared. With the aim of ensuring that equal fractions of the pressure wave were reflected at the exit of the larynx tube in both cases, the ratio of larynx tube cross-section area to maximum cross-section area in the cavity was made 0.25 in each case. The area ratios were not equal at the forward outlet of the pharynx cavity, since tongue constriction area was 0.5 cm^2 in both cases. Formant frequencies and auditory quality were appropriate for [ɑ]. F2 was too low in frequency for [i] and the auditory quality was not good. A suitable voice source was used for judging vowel quality.

GLOTTAL LOSSES IN A COMPOSITE MODEL OF SPEECH PRODUCTION

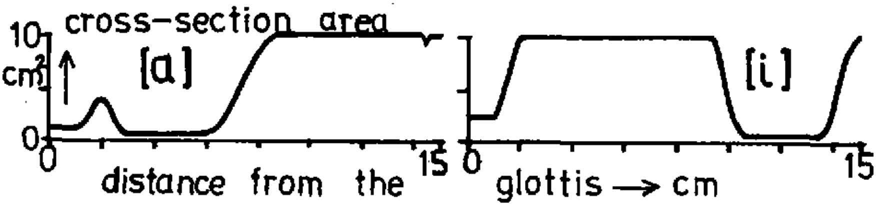

Figure 1 Two contrasting vocal tract area functions

A sine(x)/x source, flat to within +/-0.5 dB up to 5 kHz, was used; the radiation block of the model was omitted. Thus the vocal tract transfer function was studied. Two values for RCg, 0.8 and 0.4, were combined with two values for RCm, -0.8 and -0.4. Peak frequencies and bandwidths, defined as 3 dB down from the peak, were computed from an FFT program with a Hanning window. Results, interpreting spectral peaks as specific poles of a vowel-like vocal tract, are given in Table 1.

Table 1 Frequencies and bandwidths for poles of the vocal tract transfer function as vocal tract shape, glottal losses and mouth losses are changed. Frequencies in Hz to the nearest 5 Hz.
* means that a peak or one of the 3 dB down points could not be detected.

RCg	RCm	[ɑ]-type				[i]-type			
(glottal loss)	(mouth loss)	P1 (BW1)	P2 (BW2)	P3 (BW3)	P4 (BW4)	P1 (BW1)	P2 (BW2)	P3 (BW3)	P4 (BW4)
0.8 (low loss)	-0.8 (low loss)	720 (65)	1195 (80)	3300 (105)	4395 (80)	220 (20)	1965 (40)	3780 (100)	4720 (135)
0.4 (high loss)	-0.8	730 (315)	1195 (100)	3300 (110)	4440 (145)	220 (45)	1990 (110)	3905 (285)*	4755 (180)
0.8	-0.4 (high loss)	745 (100)	1100 to 1200 * (very wide?)*	3350 (435)*	4395 (120)	220 (32)	1965 (40)	3780 (100)	4860 (290)
0.4	-0.4	745 (385)	1100 to 1200 * (very wide?)*	3335 (425)*	4440 (200)	220 (55)	1990 (115)	3895 (365)	4895 (315)

The estimated accuracy of frequency measurement was +/-6 Hz approximately. Bandwidth measures seem more likely to be susceptible to error, since absence of a sample at the true peak could significantly alter the 3 dB down points.

In most cases increased losses at the glottis or mouth resulted in increased bandwidths. BW2 and BW3 for [i] were almost independent of mouth losses while BW3 for [ɑ] was almost independent of glottal losses. Under comparable loss conditions, BW1 was always greater for the [ɑ]-type vocal tract than for the [i] shape, as expected. BW2 and BW3 were greater for [ɑ] in some cases and greater for [i] in others. BW4 was always greater for [i] than for [ɑ].

Glottal and other contributions to formant bandwidths have been assessed separately using time domain (1) or frequency domain (7) transmission line models. Quantitative comparisons are difficult to make, partly because vocal tract lengths and shapes differ. BW1 for [i] and [ɑ] in Table 1 are both slightly higher than corresponding contributions of glottal loss alone shown by Flanagan et al., Figure 5 (1). BW1, BW2 and BW3 for [i] in Table 1 are lower than values given by Wakita and Fant, Table II-A-IV, where the subglottal system is included (7). Data from real speech are included in each of these studies: by Dunn (8) and by Fujimura and Lindqvist (7). Table 2 summarises the trends in formant bandwidths.

Table 2 Bandwidths of spectral peaks in the model compared with real speech data. Formant frequencies refer to the model. Frequencies are in Hz.

study	approximate formant frequency and vowel type					
	220 [i] BW1	730 [ɑ] BW1	1200 [ɑ] BW2	1980 [i] BW2	3300 [ɑ] BW3	3800 [i] BW3
our model RCG = 0.8	20-32	65-100	80 or more	40	105-435	
RCG = 0.4	45-55	315-385	100 or more	110-115	110-425	
Dunn (8) Table II	38	60	50	66	102	171
Fujimura (1) and Lindqvist: mean for females mean for males	~100 70	49 42				
Flanagan et al (1) total, Figure 4	65	80	88	120	>140	
Wakita and Fant (7) Table II-A-IV closed glottis	49	29 [a]	37 [a]	29	64 [a]	270
Table II-A-VI large glottis Ag = 0.16 cm	146	688	451	282	69	93

The range in these - and other - studies is so wide that precise guidelines for the modelling are not apparent. It seems that the modelled bandwidths are of the right order of magnitude. Losses introduced near to the middle of the vocal tract are expected to have less influence on bandwidths (7). Table 3 confirms the small effect of increasing the losses through a small opening of the velopharyngeal port.

Table 3 The effect on bandwidths of doubling the velopharyngeal port area Av. RCg = 0.8, RCm = -0.9. Frequencies to the nearest 5 Hz.

	[ɑ]-type				[i]-type			
Av cm	P1 (BW1)	P2 (BW2)	P3 (BW3)	P4 (BW4)	P1 (BW1)	P2 (BW2)	P3 (BW3)	P4 (BW4)
0.05	720 (60)	1210 (50)	3300 (60)	4395 (70)	230 (25)	1975 (40)	3780 (95)	4710 (95)
0.1	720 (60)	1210 (55)	3300 (65)	4395 (70)	230 (30)	1975 (45)	3780 (95)	4710 (95)

GLOTTAL LOSSES IN A COMPOSITE MODEL OF SPEECH PRODUCTION

CYCLICAL VARIATIONS IN GLOTTAL DAMPING

Variations in glottal damping during the voice cycle are important: oscillations need to decay before the next time of excitation of the vocal tract resonances. The pattern of decay may be a characteristic of different speaker types. As a first step to making the cyclical variations in glottal losses consistent with the cyclical variations in conditions at the glottis, an a-c component of glottal area, Ag' was obtained from the voice acoustic source Ug' as follows:

$$Ag' = \frac{\overline{Ag}}{|\overline{Ug}|} \cdot Ug'$$

This mapping does not take into account the difference between the shapes of the glottal area and glottal airflow curves, specifically the greater asymmetry in Ug' due to inertance in and above the glottis. Ag and Ag' were combined in three ways to give a total glottal area function Ag:

(1) $Ag = Ag' + \overline{Ag}$

(2) $Ag = w.Ag' + (1-w).\overline{Ag}$
where w = VOIA/max(VOIA) and VOIA is the envelope of Ug'.

(3) $Ag = \frac{\overline{Ag}}{|\overline{Ug}|} \cdot (Ug' - VOIA/2)$

Slight changes in some bandwidths were apparent on spectrograms; more detailed studies of single formant decay during the voice cycle remain to be made. The last two lines in Table 2 (7) suggest the kind of bandwidth magnitude swings which may be needed.

ACKNOWLEDGEMENTS

This research is supported by SERC grant GR/B/34874. The original version of the acoustic program was kindly supplied by colleagues at the Joint Speech Research Unit.

REFERENCES

1. J.L.FLANAGAN, K.ISHIZAKA and K.L.SHIPLEY 1974 Bell System Tech. J. 54, 485-506. Synthesis of speech from a dynamic model of the vocal cords and vocal tract.
2. J.L.KELLY and C.C.LOCHBAUM 1962 Proc. 4th Intl. Congr. on Acoustics, Copenhagen, paper G42, 1-4. Speech synthesis.
3. D.DEGRYSE 1982 Paper from ENSERG, Grenoble.. Temporal simulation of wave propogation in the lossy vocal tract.
4. E.ALLWOOD and C.SCULLY 1982 Proc. ICASSP82, Paris May 1982, Vol 2, 932-935. A composite model of speech production.
5. G.FANT 1980 Stockholm STL-QPSR 2-3/1980, 17-37. Voice source dynamics.
6. G.FANT 1979 Stockholm STL-QPSR 1/1979, 85-107. Glottal source and excitation analysis.
7. H.WAKITA and G.FANT 1978 Stockholm STL-QPSR 1/1978, 9-29. Toward a better vocal tract model.
8. H.K.DUNN 1961 J. Acoust. Soc. of Amer. 33, 1737-1746. Methods of measuring vowel formant bandwidths.